# Structural Landmarking and Interaction Modelling: A "SLIM" Network for Graph Classification

Yaokang Zhu[1], Kai Zhang[1*], Jun Wang[1*], Haibin Ling[2], Jie Zhang[3], Hongyuan Zha[4]
[1]School of Computer Science and Technology, East China Normal University, Shanghai, China
[2]Stony Brook University , New York, USA
[3] Institute of Brain-Inspired Intelligence, Fudan University, Shanghai, China
[4] School of Data Science, Shenzhen Institute of Artificial Intelligence and Robotics for Society
The Chinese University of Hong Kong, Shenzhen, China
52184501026@stu.ecnu.edu.cn,{kzhang980, wongjun, haibin.ling, jzhang080}@gmail.com, zhahy@cuhk.edu.cn

AAAI_2022

gesis Leibniz-Institut für Sozialwissenschaften

Forschungszentrum L3S Research Center

Reported by Xinsheng Wang

Artificial



Leibniz-Institut für Sozialwissenschaften

# Method

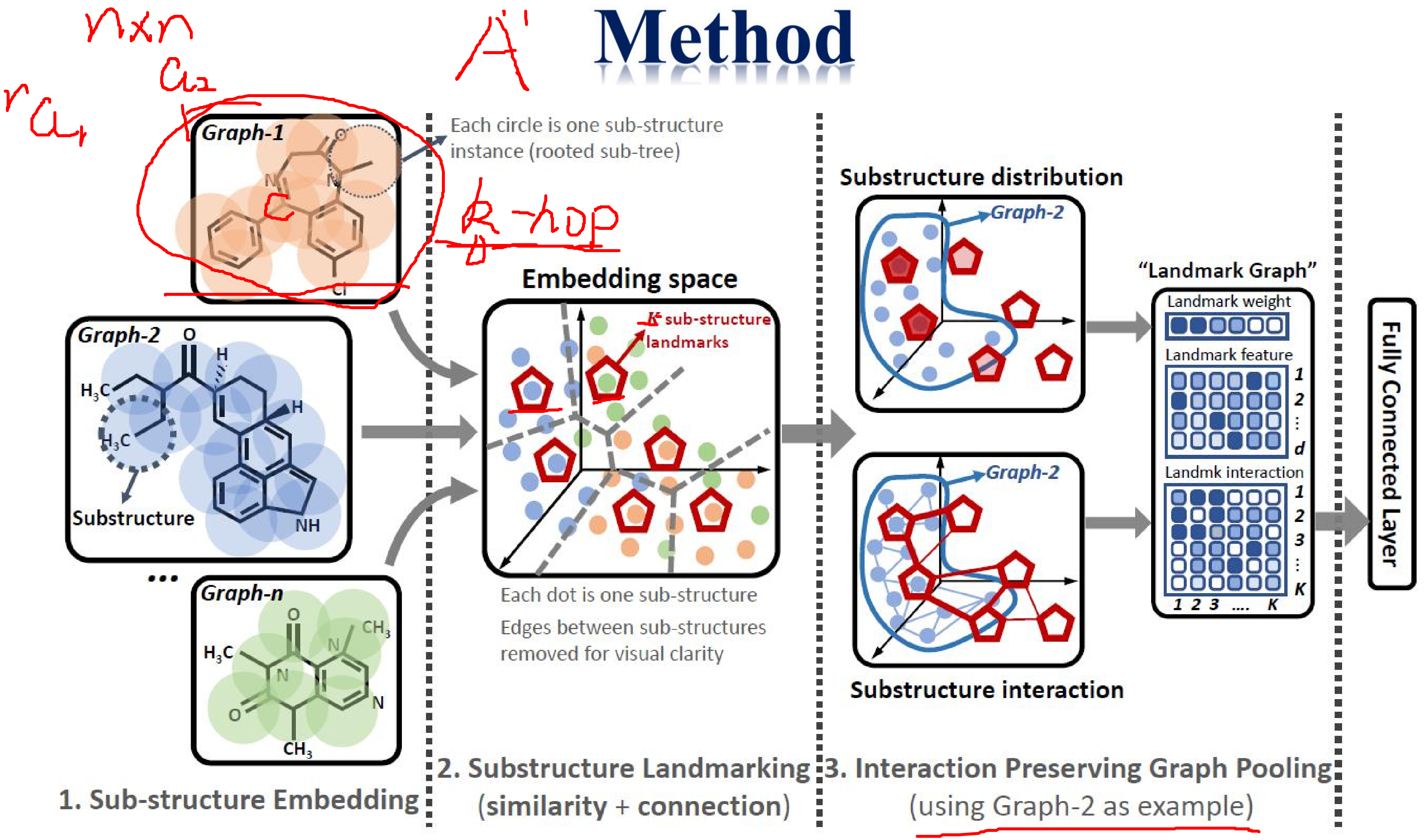



Figure 1: Three main steps of SLIM. (1) Sub-structure embedding: extract local sub-graphs and embed them in a metric space. (2) sub-structure landmarking: compute sub-structure representatives through unsupervised clustering across graphs. (3) Identity-preserving graph pooling: project each graph on the common set of sub-structure landmarks for final prediction.

# Method

## 2-hop Sub-structure Identification and Embe...

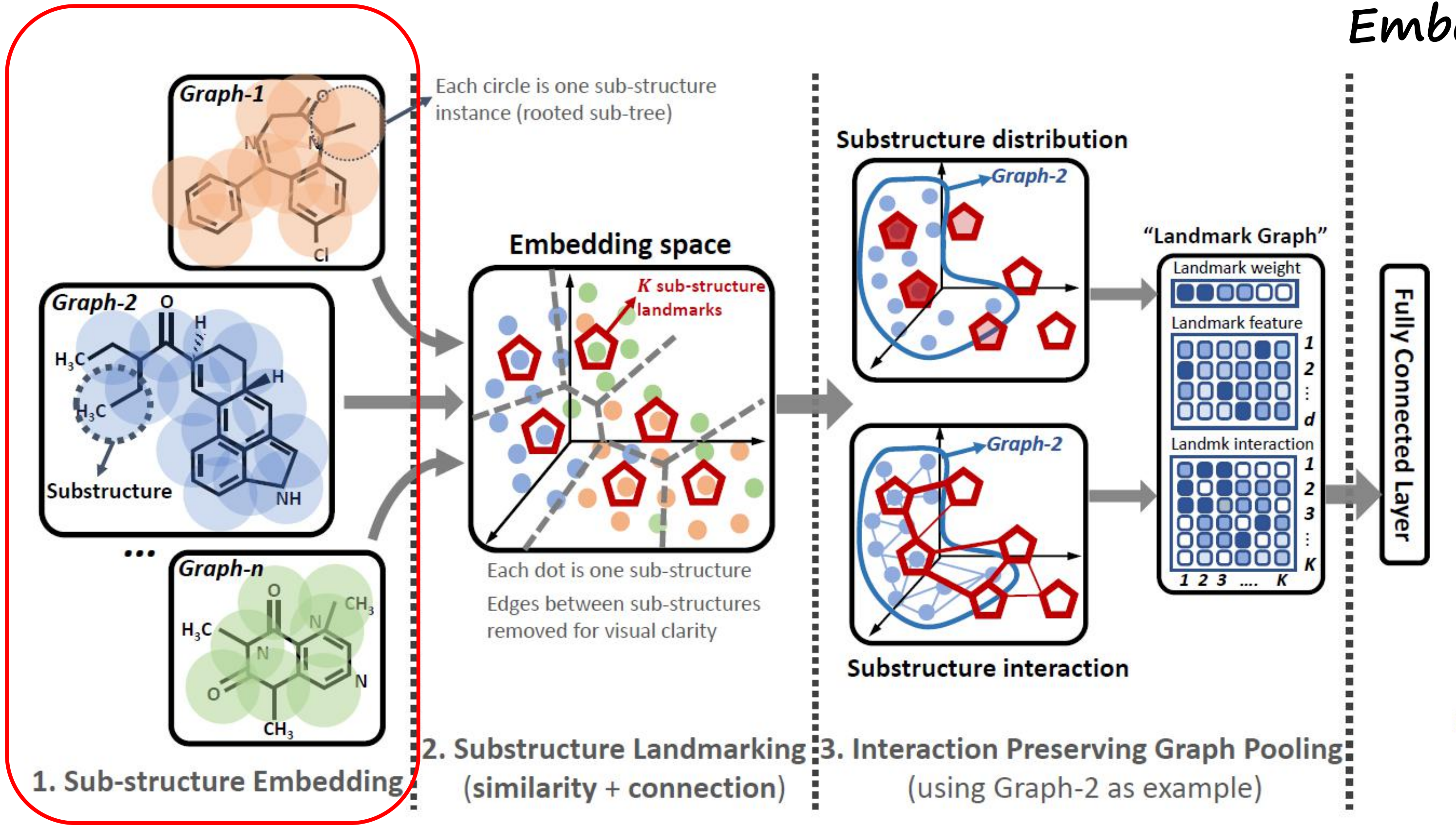



Figure 1: Three main steps of SLIM. (1) Sub-structure embedding: extract local sub-graphs and embed them in a metric space. (2) sub-structure landmarking: compute sub-structure representatives through unsupervised clustering across graphs. (3) Identity-preserving graph pooling: project each graph on the common set of sub-structure landmarks for final prediction.

$$\mathbf{Z}_i = \mathbf{A}_i^{(k)} \mathbf{X}_i \quad (1)$$

$$\mathbf{A}_i^{(k)} = \mathbf{I} + \tilde{\mathbf{A}}_i^{(1)} + \tilde{\mathbf{A}}_i^{(2)} \ldots + \tilde{\mathbf{A}}_i^{(k)}$$

$$\mathbf{Z}_i = [\mathbf{X}_i \ \tilde{\mathbf{A}}_i^{(1)}\mathbf{X}_i \ \tilde{\mathbf{A}}_i^{(2)}\mathbf{X}_i \ \ldots \ \tilde{\mathbf{A}}_i^{(k)}\mathbf{X}_i], \quad (2)$$

namely $\tilde{\mathbf{A}}_i^{(k)}(p,q) = 1$ if node $p$ and $q$ are *exactly* $k$-hops away in graph $\mathcal{G}_i$, and 0 otherwise.

$$f(\mathbf{Z}_i) = \sigma\left(\mathbf{Z}_i \cdot \mathbf{T} + \mathbf{B}\right), \quad (3)$$

where $\mathbf{T}$ is transform matrix, $\mathbf{B}$ is bias matrix (a bias vector repeated $n_i$ times row-wise) and $\sigma(\cdot)$ is the RELU function.

$$\mathbf{H}_i = f(\mathbf{Z}_i)$$

$$\max \sum_{i=1}^{n} \sum_{j=1}^{n_i} \sum_{l \in \mathcal{N}_j^i} \log \left( \frac{\exp\langle \mathbf{H}_i(j,:), \mathbf{H}_i(l,:)\rangle}{\sum_{l'} \exp\langle \mathbf{H}_i(j,:), \mathbf{H}_i(l',:)\rangle} \right). \quad (4)$$

L'表示第i个图中的所有节点

Here $\mathbf{H}_i(j,:)$ is the $j^{th}$ row of $\mathbf{H}_i$, $\langle,\rangle$ is inner product, and $\mathcal{N}_j^i$ are the neighbors of node $i$ in graph $\mathcal{G}_i$.

# Method

Sub-structure Landmarking

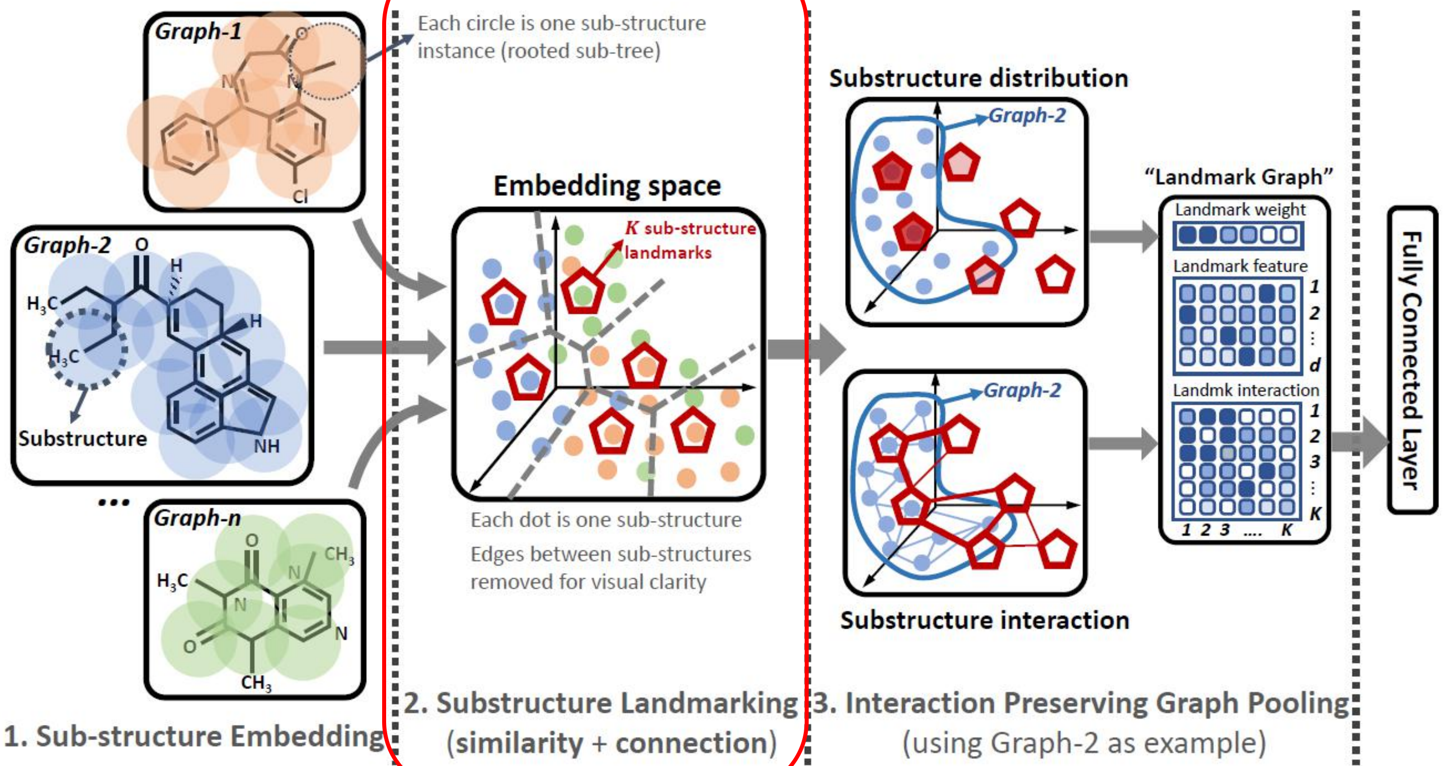



Figure 1: Three main steps of SLIM. (1) Sub-structure embedding: extract local sub-graphs and embed them in a metric space. (2) sub-structure landmarking: compute sub-structure representatives through unsupervised clustering across graphs. (3) Identity-preserving graph pooling: project each graph on the common set of sub-structure landmarks for final prediction.

kc

Let $\mathbf{U} = \{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, ..., \boldsymbol{\mu}_K\}$ be structural landmarks in the latent space of embedded sub-structures. To fully represent diverse sub-structures, each sub-graph instance should be faithfully approximated by the closest landmark. We use a soft assignment matrix $\mathbf{W}_i \in \mathbb{R}^{n_i \times k}$ for each graph $\mathcal{G}_i$, whose $jk^{th}$ entry is the probability that the $j^{th}$ sub-structure from $\mathcal{G}_i$ belongs to the $k^{th}$ landmark $\boldsymbol{\mu}_k$. Inspired by the deep embedding clustering (Junyuan, Ross, and Ali 2016), $\mathbf{W}_i$ is parameterized by a Student's t-distribution

$$\mathbf{W}_i(j,k) = \frac{(1 + \|\mathbf{H}_i(j,:) - \boldsymbol{\mu}_k\|^2/\alpha)^{-\frac{\alpha+1}{2}}}{\sum_{k'}(1 + \|\mathbf{H}_i(j,:) - \boldsymbol{\mu}_{k'}\|^2/\alpha)^{-\frac{\alpha+1}{2}}}, \quad (5)$$

use $\alpha = 1$

$$\min_{\mathbf{U}, \mathbf{H}_i' s} \sum_i \mathrm{KL}\left(\mathbf{W}_i, \widetilde{\mathbf{W}}_i\right)$$

$$s.t. \quad \widetilde{\mathbf{W}}_i(j,k) = \frac{\mathbf{W}_i^2(j,k)/\sum_l \mathbf{W}_i(l,k)}{\sum_{k'}\left[\mathbf{W}_i^2(j,k')/\sum_l \mathbf{W}_i(l,k')\right]}. \quad (6)$$

$$q_{ij} = \frac{(1 + \|z_i - \mu_j\|_2^2/\alpha)^{-\frac{\alpha+1}{2}}}{\sum_{j'}(1 + \|z_i - \mu_{j'}\|_2^2/\alpha)^{-\frac{\alpha+1}{2}}} \quad (2)$$

reedom of the Student's $t$-distribution, which is set to 1 as suggested

rget distribution $p_i$ is derived by manipulating the obtained soft ngthen high confidence predictions. It can be formalized as follows:

$$p_{ij} = \frac{q_{ij}^2/\sum_i q_{ij}}{\sum_{j'}(q_{ij'}^2/\sum_i q_{ij'})} \quad (3)$$

# Method

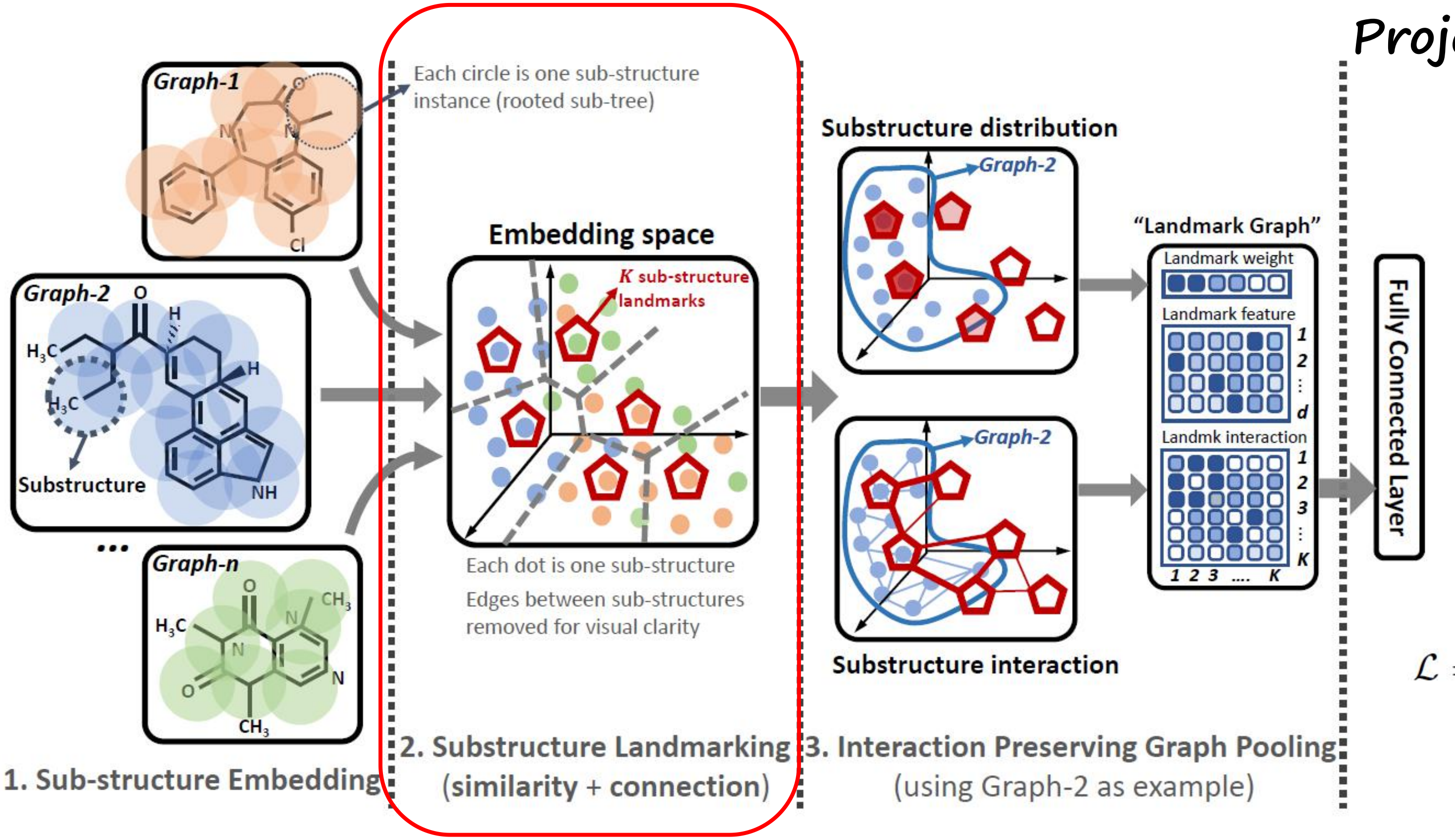



Figure 1: Three main steps of SLIM. (1) Sub-structure embedding: extract local sub-graphs and embed them in a metric space. (2) sub-structure landmarking: compute sub-structure representatives through unsupervised clustering across graphs. (3) Identity-preserving graph pooling: project each graph on the common set of sub-structure landmarks for final prediction.

## Interaction Preserving Graph Projection

Landmark weight

$$\mathbf{p}_i = \mathbf{W}_i' \cdot \mathbf{1}_{n_i \times 1}. \tag{7}$$

Landmark feature

$$\mathbf{M}_i = \mathbf{X}_i' \cdot \mathbf{W}_i \cdot \mathbf{P}_i^{-1} \tag{8}$$

$$\mathbf{P}_i = \mathrm{diag}(\mathbf{p}_i)$$

Landmark Interaction

$$\mathbf{C}_i = \mathbf{W}_i \cdot \mathbf{A}_i \cdot \mathbf{W}_i', \tag{9}$$

$$\tilde{\mathbf{C}}_i = \mathbf{P}_i^{-1} \mathbf{C}_i \mathbf{P}_i^{-1}$$

$$\mathcal{L} = \underbrace{-\lambda_1 \cdot \mathcal{L}_{pn2v} + \lambda_2 \cdot \mathcal{L}_{deep-clustering}}_{regularization-terms} + \underbrace{\mathcal{L}_{cross-entropy}}_{classification-loss}$$

# Method

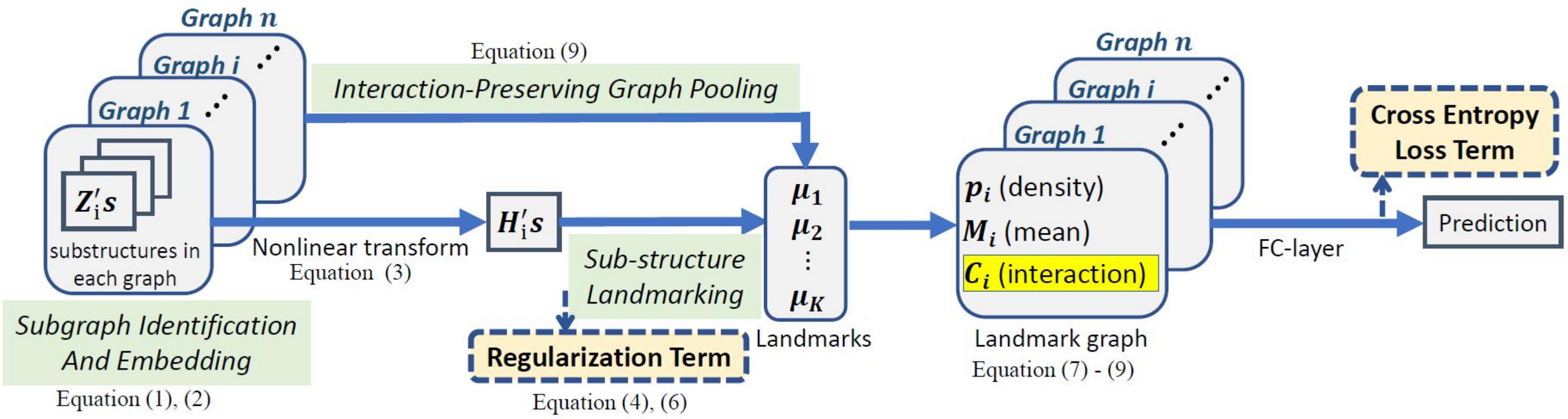



Figure 2: End-to-end training architecture of the SLIM network.

$$\mathcal{L}_{pn2v} \quad \max \sum_{i=1}^{n} \sum_{j=1}^{n_i} \sum_{l \in \mathcal{N}_j^i} \log \left( \frac{\exp\langle \mathbf{H}_i(j,:), \mathbf{H}_i(l,:)\rangle}{\sum_{l'} \exp\langle \mathbf{H}_i(j,:), \mathbf{H}_i(l',:)\rangle} \right). \quad (4)$$

$$\mathcal{L} = \underbrace{-\lambda_1 \cdot \mathcal{L}_{pn2v} + \lambda_2 \cdot \mathcal{L}_{deep-clustering}}_{regularization-terms} + \underbrace{\mathcal{L}_{cross-entropy}}_{classification-loss}$$

$$\mathcal{L}_{deep_clustering} \quad \min_{\mathbf{U}, \mathbf{H}_i's} \sum_i \mathrm{KL}\left(\mathbf{W}_i, \widetilde{\mathbf{W}}_i\right)$$

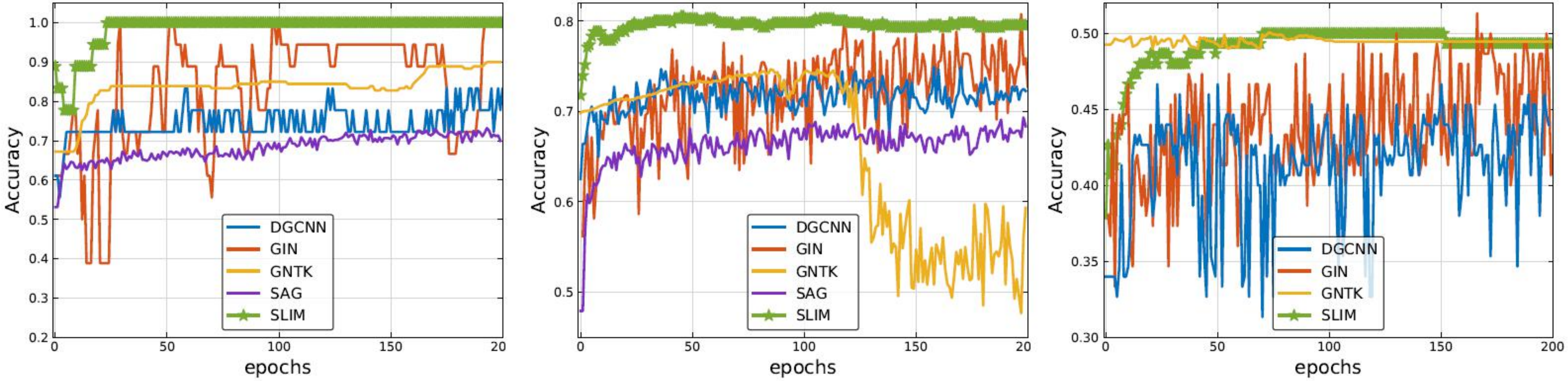
$$s.t.\ \widetilde{\mathbf{W}}_i(j,k) = \frac{\mathbf{W}_i^2(j,k) / \sum_l \mathbf{W}_i(l,k)}{\sum_{k'} \left[\mathbf{W}_i^2(j,k') / \sum_l \mathbf{W}_i(l,k')\right]}. \quad (6)$$

# Experiments

Table 1: Classification on benchmark data-sets (cheminformatics, bioinformatics & social networks).

| ALG. | MUTAG | PTC | NCI1 | Protein | D&D | IMDB-B | IMDB-M | COLLAB |
|---|---|---|---|---|---|---|---|---|
| GK | 81.38±1.74 | 55.65±0.46 | 62.49±0.27 | 71.39±0.31 | 74.38±0.69 | 65.87±0.98 | 43.89±0.38 | 72.84±0.28 |
| PK | 76.00±2.69 | 59.50±2.44 | 82.54±0.47 | 73.68±0.68 | 78.25±0.51 | — | — | — |
| WLGK | 84.11±1.91 | 57.97±2.49 | **84.46±0.45** | 74.68±0.49 | 78.34±0.62 | 73.40±4.63 | 49.33±4.75 | 79.02±1.77 |
| PC-SAN | 92.63±4.21 | 60.00±4.82 | 78.59±1.89 | 75.89±2.76 | 77.12±2.41 | 71.00±2.29 | 45.23±2.84 | 72.60±2.15 |
| DGCNN | 85.83±1.66 | 58.59±2.47 | 74.46±0.47 | 75.54±0.94 | 79.37±1.03 | 70.03±0.86 | 47.83±0.85 | 73.76±0.49 |
| DiffPool | 90.52±3.98 | — | 76.53±2.23 | 75.82±3.56 | 78.95±2.40 | 73.58±3.24 | 52.13±2.71 | 79.70±1.84 |
| GNTK | 90.12±8.58 | 67.92±6.98 | 75.20±1.53 | 75.61±4.24 | 79.42±2.18 | 75.93±3.61 | 52.82±4.65 | **83.60±1.22** |
| SAG | 73.53±9.68 | 69.67±3.12 | 74.18±1.29 | 71.86±0.97 | 76.91±2.12 | 72.61±2.23 | 51.80±2.08 | 79.88±1.02 |
| GIN | 90.03±8.82 | 64.60±7.00 | 79.84±4.57 | 75.28±2.65 | 77.58±2.94 | 75.15±5.08 | 52.33±2.84 | 80.21±1.92 |
| StrPool | 82.21±3.13 | 71.46±2.21 | 71.31±1.14 | 76.89±1.67 | **79.72±1.98** | 73.77±2.01 | 50.17±1.28 | 79.14±0.88 |
| SLIM | **93.28±3.36** | **72.41±6.92** | **80.53±2.01** | **77.47±4.34** | 79.61±2.66 | **77.23±2.12** | **53.38±4.02** | 78.22±2.02 |

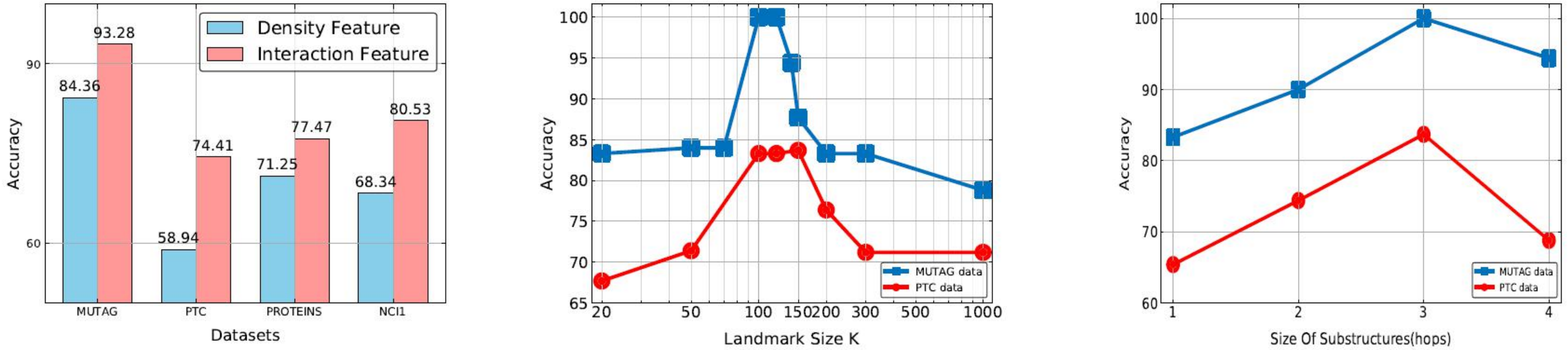
# Experiments




(a) MUTAG data. (b) NCI data. (c) IMDB-M data.

Figure 3: SLIM has a stable performance based on the accuracy-vs-epoch curve.

# Experiments




(a) Interaction vs density (counting). (b) Impact of landmark set size $K$. (c) Impact of sub-structure size.

Figure 4: The performance of SLIM w.r.t. the choice of hyper-parameters and graph level feature.

In Figure 4(a), we compare performance of SLIM when using the weights (or density) of the landmark $\mathbf{p}_i$ (7), or the interaction matrix $\mathbf{C}_i$ (9), as graph-level features. The interaction feature consistently generates better accuracy than distribution-based features, validating the importance of modeling the interacting relation in graph classification tasks.

# Thank you!